DOCUMENT RESUME

ED 291 783 TM 011 068

AUTHOR Cason, Gerald J.; And Others
TITLE Integrated Test Scoring, Performance Rating and Assessment Records Keeping.
PUB DATE Nov 87
NOTE 20p.; Paper presented at the Annual Meeting of the Association of Medical Colleges (Washington, DC, November 6-12, 1987).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Clinical Experience; *Computer Assisted Testing; Educational Assessment; Essay Tests; Graduate Medical Students; Interrater Reliability; Item Analysis; Medical Education; Multiple Choice Tests; Standardized Tests; Test Reliability; *Test Scoring Machines; Weighted Scores
IDENTIFIERS FORTRAN Programing Language; *Objective Test Scoring Performance Rating

ABSTRACT

The Objective Test Scoring and Performance Rating (OTS-PR) system is a fully integrated set of 70 modular FORTRAN programs run on a VAX-8530 computer. Even with no knowledge of computers, the user can implement OTS-PR to score multiple-choice tests, include scores from external sources such as hand-scored essays or scores from nationally administered examinations, use and revise multiple different performance rating inventories in the same or different course, assign variable points to test questions and rating inventory items, assign variable weight to different assessments, maintain a grade book linking student names and identification numbers to scores on multiple assessments, and produce reports on student performance and on the performance of the assessment procedure itself (including item analysis, interrater reliability, test reliability). In an average month, OTS-PR supports 50 courses and other applications in 4 colleges of the University of Arkansas for Medical Sciences. Use of the more valid National Board Examinations Type X question is replacing Type K items. Most junior year clerkships and two residencies use the OTS-PR system to rate clinical performance. Use of essay questions is increasing. Samples are provided of student enrollment sheets, test answer sheets, and various reports and inventories. (TJH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED291783

INTEGRATED TEST SCORING, PERFORMANCE RATING AND ASSESSMENT RECORDS KEEPING[1]

Gerald J. Cason, PhD
Office of Educational Development
System Designer and Developer

Carolyn L. Cason, RN, PhD
Graduate Program, College of Nursing
Associate System Designer

Ture W. Schoultz, PhD
College of Medicine
System Development Sponsor

UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES

Little Rock, Arkansas 72205

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Gerald Carson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

TM011 068

Correspondence to:
Gerald J. Cason, PhD
UAMS-OED-595
4301 West Markham Street
Little Rock, AR 72205

Telephone: (501) 661-5720

[1]Paper presented as part of an exhibit in Innovations in Medical Education, Association of American Medical Colleges, Washington, D.C., November 6-12, 1987

BEST COPY AVAILABLE

INTEGRATED TEST SCORING, PERFORMANCE RATING AND ASSESSMENT RECORDS KEEPING[2]

G.J. Cason, Ph.D., C.L. Cason, Ph.D. & T.W. Schoultz, Ph.D.
University of Arkansas for Medical Sciences
Little Rock, AR 72205; (501) 661-5720

Objectives. Develop a system for processing a wide variety of student performance assessment data (e.g., exams, essay scores, clinical performance ratings), maintenance of this data in an integrated course level data base, and generation of appropriate reports. The system should require no computer expertise on the user's part, be easy to use, and meet the needs of both Basic Science and Clinical course faculty.

Methods. The UAMS Objective Test Scoring and Performance Rating (OTS-PR) system is a fully integrated set of 70 modular FORTRAN programs run on a VAX-8530. The user needs no knowledge of computers because the user interacts with a full-time Assessment Scoring Coordinator (ASC) who in turn interacts with the system hardware and software. Although the ASC enters some data through a terminal keyboard (e.g., instructor's name, course name, etc.) most data enter the system via an NCS-7008 scanner and scannable data sheets (e.g., student enrollment, test answer and performance rating sheets). The computer's line printer pre-prints students' names and ID numbers (in both ordinary characters and in scannable form) on test answer sheets and rating sheets. It is also used to print easily customized and revised sets of rating items (criteria) on the rating sheets.

OTS-PR makes it easy for the user (i.e., course director or coordinator) to: score multiple choice exams, including questions requiring more than a single mark on the answer sheet such as NBME Item Type X; include scores from external sources such as hand-scored essays or scores from nationally administered examinations; use and quickly revise multiple different performance rating inventories in the same and/or different course(s) or clerkship(s); assign variable points to test questions and rating inventory items (criteria); assign variable weight to different assessments (tests, ratings); maintain a gradebook linking student names and IDs to their scores on multiple assessments; and, produce appropriate reports on students' performance, on the current assessment or cumulatively; and reports on the performance of the assessment procedure itself (e.g., item analyses, inter-rater or test reliability).

Results. In an average month OTS-PR supports 50 courses and other applications in the four UAMS Colleges. In the College of Medicine, all the Basic Science courses use OTS. Use of the more valid NBME Type X question is gradually replacing the less valid Type K. Almost all courses have student IDs pre-printed on answer sheets thus eliminating problems with students not remembering or improperly entering ID information on the sheets. Most Junior year Clerkships and two residencies use the OTS-PR system to collect and process ratings of learners' clinical performance. Each course uses rating criteria chosen to fit its individual needs. Use of essay questions is increasing.

Conclusions. Although many enhancements are still needed, for example being able to automatically off-set differences in raters' standards (i.e., rater bias) in clinical courses, availability of OTS-PR often makes it easy to use an appropriate assessment approach and thus encourages faculty to do so. When faculty are given student performance assessment tools, such as OTS-PR, that make the task easy and convenient, the quality control of medical education improves.

[2]Abstract as published in proceedings of 1987 AAMC-IME.

INTEGRATED TEST SCORING, PERFORMANCE RATING AND ASSESSMENT RECORDS KEEPING

G.J. Cason, Ph.D., C.L. Cason, Ph.D. & T.W. Schoultz, Ph.D.
University of Arkansas for Medical Sciences
Little Rock, AR 72205; (501) 661-5720

OVERVIEW: The UAMS Objective Test Scoring and Performance Rating (OTS-PR) system is described. OTS-PR provides scoring, records keeping and reporting for a wide variety of performance assessment data. User access is eased by the Assessment Scoring Coordinator (ASC) who communicates with the main-frame (VAX-8530) computer, data sheet optical scanner, etc. The system scores multiple-choice exams, including question types requiring more than one mark to indicate the correct answer(s) to a single question, such as NBME Item Type X. Scores from manually scored exercises (e.g., essays, reports) and external sources (e.g., nationally administered examinations) may be easily included. Special provisions, which permit great flexibility within and between courses, are made for clinical performance ratings. All assessment data are maintained in an integrated course-level data base. Included here are examples of many of the wide variety of reports that are available from OTS-PR for individual and multiple assessments; for examinees and raters; and, for analyses of entire assessments and individual test or rating items. Reports had by a user are selected from a menu of those available. Reports are normally available on a next-day basis. The system is equally useful to Basic Science and Clinical faculty. The ease of use, flexibility, accuracy, and speed of the system encourage faculty to adopt and use more effective and appropriate assessment methods.

Description of the UAMS OTS-PR System

Figure 1 provides an overview of information flow in the UAMS Objective Test Scoring and Performance Rating (OTS-PR) system. At the left center is the Assessment Scoring Coordinator (ASC) in the Department of Computer Services' Academic Support Center. The User (i.e., Course Director or Secretary), located at the right center of Figure 1, interacts with the ASC, who in turn interacts with the sheet scanner and the 70 modular FORTRAN programs composing OTS-PR which run on a main-frame computer. Currently, the system uses a National Computer System 7008 trans-optic data sheet scanner. Prior to 1987, the system was implemented on a Digital Equipment Corporation DEC-10, it has now been converted to run on a VAX-8530. Information is passed from user to ASC in three media: orally; written request forms (which are processed manually); and, scan (i.e., machine scannable data) sheets. Table 1 lists the information (variable) types a user may maintain in an OTS-PR course level data base, along with the maximum number of cases/data elements for each. Table 2 provides a descriptive listing of the request forms and scan sheets used in OTS-PR. Table 3 gives a partial listing of the 17 currently available OTS-PR reports.

Figure 1. Over-view of Information Flow in the UAMS OTS-PR System

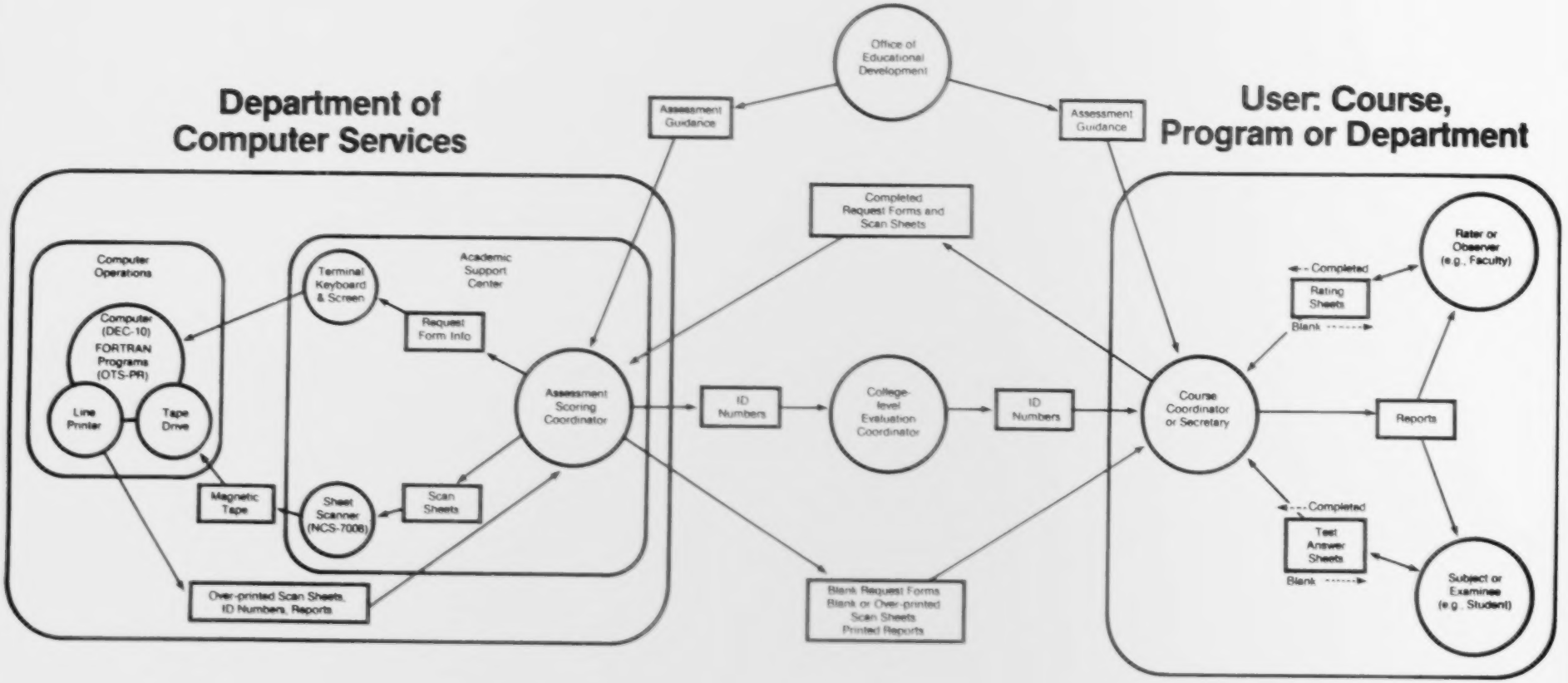


Table 1. Current Size Limitations in OTS-PR

Maxima	Feature (Variable)
200	Subject (e.g., student) names and ID numbers
200	Rater (e.g., faculty) names and ID numbers
30	Assessments (Tests, Ratings; Settings, Occasions) per course
12	Categories (sub-scales) per assessment (test or rating inventory)
30	Rating inventories (item/criteria sets) per course
40	Rating items (criteria) per inventory
240	Test items (questions) per test
5	Test Forms: Alternative orders of questions in same test

Table 2. Request Forms and Scan Sheets Used in OTS-PR

Manual Input via Request Forms
- New User (Start-up: info for report header, etc.)
- Report and Gradebook Selection
- Rating Inventory Specification (Scale definitions, item text)
- Scan Sheet Over-printing
- Gradebook Assessment Weight Change
- Assessment (Test, Rating) Scoring

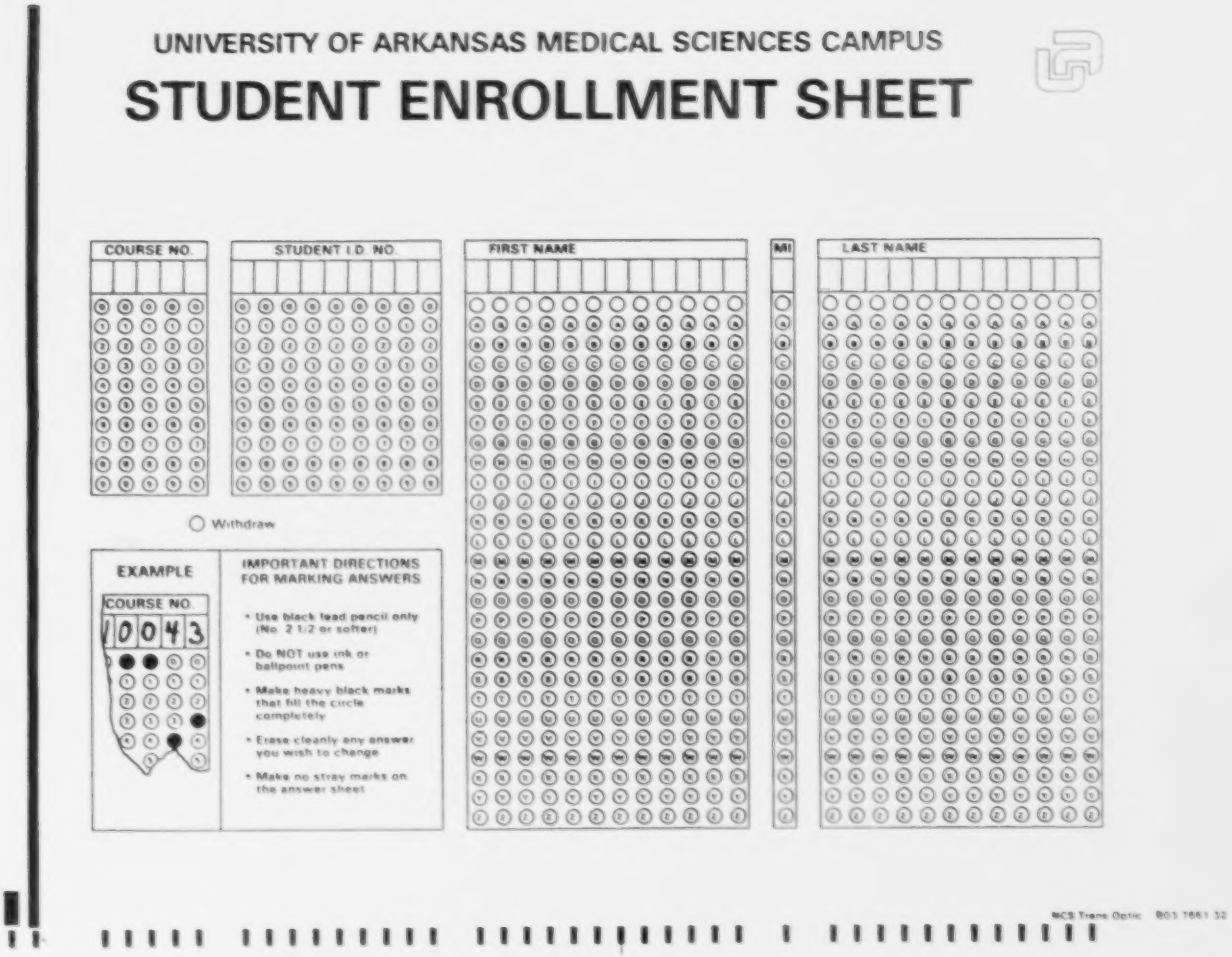
Automated Input via Scan Sheets
- Student/Rater Enrollment
- Rating (Must be slug printed)
- Test Answer (May be slug printed)
- Control: Alternative Test Forms (Must be slug printed)

Table 3. Partial Listing of OTS-PR Reports

For Current Assessment (Test or Rating)
- Assessment Instrument Analysis
  - Analysis Summary
  - Item Analysis
- Subject (e.g., student) Performance
  - Individual Performance Report (Item level ratings)
  - Letter Report (Item level test performance)
- Group Reports
  - Department (Total and category scores; for archive)
  - Subjects by Rank Order
  - Histogram of Subject Scores
  - Posting (Subjects' names excluded)
- Rater Performance (Reports parallel those for Subjects)

Across Multiple Assessments (Tests, Ratings; Settings, Occasions)
- For Course Coordinator's Use
  - Alpha ordered, All Subjects' Scores to Date
  - Rank ordered, Subjects' Cumulative Weighted Totals
- Posting (Subjects' names excluded, ID ordered)
  - All Scores to Date
  - Cumulative Weighted Total

Figure 2. Student Enrollment Sheet


UNIVERSITY OF ARKANSAS MEDICAL SCIENCES CAMPUS
STUDENT ENROLLMENT SHEET
COURSE NO.
STUDENT I.D. NO.
FIRST NAME
MI
LAST NAME
Withdraw
EXAMPLE
COURSE NO.
1043
IMPORTANT DIRECTIONS FOR MARKING ANSWERS
• Use black lead pencil only (No. 2 1/2 or softer)
• Do NOT use ink or ballpoint pens
• Make heavy black marks that fill the circle completely
• Erase cleanly any answer you wish to change
• Make no stray marks on the answer sheet

STUDENT ENROLLMENT SHEET (Figure 2): Provides initial optically scanned input of student names and identification (ID) numbers. Thereafter, a master class roll is maintained on computer disk. Student rolls for specific courses are created by the Assessment Scoring Coordinator (ASC) as needed. The same sheet is used to create a roll of rater names and IDs for a course in which scannable rating sheets (Figure 6) are used to record performance ratings and reports on the raters are desired.

CONFIDENTIAL IDENTIFICATION NUMBERS are used for subjects (e.g., students) and raters. There is a mathematical relationship between the left-most and right-most digits which is used by the sheet scanner to detect errors in manually entered IDs and sort those sheets into a separate bin. Immediate correction of these errors can save not only computer time, but save turn around for the final, complete reports. Batches of ID numbers are provided by the Assessment Scoring Coordinator to a designated Evaluation Coordinator in each UAMS College (usually a secretary in the Dean's Office). Each college has its own procedure for assigning ID numbers to people. If a person's ID becomes compromised, it can be easily changed.

Figure 3. Test Answer Sheets with Over-printed ID Data

THIS AREA CONTAINS NON MACHINE-READABLE IDENTIFYING INFORMATION. FILL OUT ANY MISSING INFORMATION.
STUDENT NAME: POINTS SHEET TEST NAME: COMPREHENSIVE EXAMINATION

THIS AREA WILL BE MACHINE SCANNED. FILL OUT ONLY IF SO INSTRUCTED.

THIS AREA CONTAINS NON MACHINE-READABLE IDENTIFYING INFORMATION. FILL OUT ANY MISSING INFORMATION.
STUDENT NAME: ITEM SHEET TEST NAME: COMPREHENSIVE EXAMINATION

THIS AREA WILL BE MACHINE SCANNED. FILL OUT ONLY IF SO INSTRUCTED.

THIS AREA CONTAINS NON MACHINE-READABLE IDENTIFYING INFORMATION. FILL OUT ANY MISSING INFORMATION.
STUDENT NAME: CATEGORY SHEET TEST NAME: COMPREHENSIVE EXAMINATION

THIS AREA WILL BE MACHINE SCANNED. FILL OUT ONLY IF SO INSTRUCTED.

THIS AREA CONTAINS NON MACHINE-READABLE IDENTIFYING INFORMATION. FILL OUT ANY MISSING INFORMATION.
STUDENT NAME: KEY SHEET TEST NAME: COMPREHENSIVE EXAMINATION

THIS AREA WILL BE MACHINE SCANNED. FILL OUT ONLY IF SO INSTRUCTED.

THIS AREA CONTAINS NON MACHINE-READABLE IDENTIFYING INFORMATION. FILL OUT ANY MISSING INFORMATION.
STUDENT NAME: KILDARE, NOEL C. TEST NAME: COMPREHENSIVE EXAMINATION
CLASS: INSTRUCTOR: GJ CASON, PhD
COURSE: GENERIC CLINICAL PRACTICUM (SYNTHET DATE:

THIS AREA WILL BE MACHINE SCANNED. FILL OUT ONLY IF SO INSTRUCTED.

STUDENT I.D. NUMBER: 0 8 6 2 1 2 8 2 6

STUDENT INITIALS F M L: N C K

COURSE NUMBER: 2 2 3 0 1

TEST FORM: 4

ENTER SCORE (FOR MANUAL SCORING)

THIS AREA WILL BE MACHINE SCANNED. FILL OUT ONLY IF SO INSTRUCTED.

STUDENT I.D. NUMBER: 9 9 9 9 9 9 9 9 8

STUDENT INITIALS F M L: K E Y

COURSE NUMBER: 2 2 3 0 1

TEST FORM: 4

ENTER SCORE (FOR MANUAL SCORING)

Table 4. List of Currently Available Test Item Types in OTS-PR

0: Single mark, multiple choice. Correct answer indicated with one mark.

1: Multiple mark, multiple choice. Correct answer(s) indicated with 0-5 marks. Partial credit scoring; for each correctly marked alternative 1/5 of item's total points is awarded.

2: First item in all-or-none series. Correct answer indicated with one mark.
3: Continuation of all-or-none series. Correct answer indicated with one mark. All items in an all-or-none series must be answered correctly or no credit is given for any correct answers.

4: Bonus item; all examinees receive credit regardless of their response(s).
5: Invalid item; no examinee receives credit regardless of his or her response(s). Types 4 and 5 provide alternative ways to "discard" bad items when re-scoring test following test/item analysis, student appeals, and post test critique/discussion session.

6: Alternative responses; credit given if only one of correct responses is marked. Used in re-scoring a Type 0 item that inadvertently turns out to have two or more correct responses. Because instructions are to make only a single mark, no credit given if two or more of the alternatives are marked.

7: Exact match multiple-mark multiple choice item. Can be used for dichotomous scoring of NBME Type X, but best used for matching type question where the two "lists" to be matched contain up to 30 things.

8: Likert scaled item with blank = 0; A = 20; B = 40; C = 60; D = 80; E = 100.
9: Likert scaled item with blank = 0; A = 100; B = 80; C = 60; D = 40; E = 20.

TEST ANSWER SHEET (Figure 3): Provides automated input of answers to multiple-choice questions, scores from external sources, and control information used in scoring and reporting. Scoring control data input via answer sheets with special codes in the "initials" and "Student ID Number" data grids include: (a) POINTS for each item (0-30); (b) ITEM response format TYPE (10 available listed in Table 4); (c) CATEGORY or sub-scale of each item (1-12); (d) KEY to correct answer(s) for each item. On the KEY sheet the maximum score for a manually scored exercise or score from an external source that is to be included as part (or all) of the score for this test is entered in the "manual scoring" data grid.

Errors in manual entry of ID information are eliminated by using the computer's printer to over-print ID data, including "slugging" data in the machine scannable ID grids on answer sheets for each examinee. Almost all course directors choose to use this option. The student ID number printed on the answer sheet by the computer is not confidential. Thus, the sheets require no special handling prior to answers being recorded on them. A different, confidential ID number is used in OTS-PR records and reports. It is this confidential ID number that the student enters in the ID grid if manual entry is used. Sheets with pre-slugged ID's can be mixed with ones having ID's entered manually.

A score from an externally scored exam or manually scored exercise is input through the "manual scoring" data grid on the back of the answer sheet. The system permits multiple marks for a single item or question to be valid. This allows the use of question response formats such as required by NBME Item Type X (Hubbard, 1971) which is more valid than the single mark format of NBME Item Type K (Albanese, Kent & Whitney,

1979; Harasym, Norris & Lordschieder, 1980; Whitley, Jones & Revicki, 1986; Cason, 1986). OTS-PR item type 1 is equivalent to NBME type X. Table 4 provides a descriptive listing of the ten item types currently available to OTS-PR users.

GRADEBOOKS: Provide records keeping, summation, and reporting of scores from multiple assessments (tests, ratings, etc.). The user may choose to keep Gradebook records in any one of three units of measurement: Raw Score (original points), Percent Score, or Z-score (standard scores with mean = 500; sd = 100). Explicit weighting is permitted only in the Percent and Z-score gradebooks. The Z-score gradebook is especially appropriate when scores from different measurement methods are combined, such as when performance ratings are combined with objectively scored tests.

Figure 4. Gradebook Reports for Instructor or Course Coordinator

Z-Score GRADEBOOK: Assessment List

Prepared 21-SEP-87 07:49
by the UAMS OTS/PR System (version V1)
as implemented at UAMS

Instructor: GJ CASON, PHD
Course : GENERIC CLINICAL PRACTICUM (SYNTHETIC DATA) 1985-86
Course No.: 22301

Dept: Educational Development
Slot: 595 Phone: 661-5720
Number of students currently enrolled: 128

Z-SCORES

Test No.	Date Given	Test Name	Test Weight
1	19-Jul-86	Outpatient Rating - Supervisor	1
2	19-Jul-86	Emergency Room Rating - Staff	1
3	29-Jul-86	Inpatient Rating - Faculty	1
4	29-Jul-86	Comprehensive Examination	3
		Total:	6

Z-Score GRADEBOOK: Instructor's Alpha Ordered Cumulative Score Report Prepared 21-SEP-87 07:49

Z-SCORES

ID	Name	Test:	1	2	3	4	Weighted Total	Total Weights	Avg *
		Weight:	1	1	1	3			
286001645	ARNOLD, FRANKLIN D.		535	ABSENT	280	497	2306	5	461.20
286007980	BEAKER, JULIO M.		ABSENT	580	721	558	2975	5	595.00
286000362	FOOT, DONNA D.		ABSENT	ABSENT	ABSENT	ABSENT	N/A	N/A	N/A
286002197	KILDARE, NOEL C.		624	ABSENT	544	298	2062	5	412.40
286001297	ZEIMANOVICH, MARINA K.		538	ABSENT	799	619	3194	5	638.80

Z-Score GRADEBOOK: Instructor's Students By Cumulative Rank Report Prepared 21-SEP-87 07:49

Z-SCORES

	ID	Avg	Rank	%ile	Final Z	Name
4	286001297	638.80	4	96	701	ZEIMANOVICH, MARINA K.
12	286007980	595.00	12	90	638	BEAKER, JULIO M.
93	286001645	461.20	93	23	445	ARNOLD, FRANKLIN D.
110	286002197	412.40	110	9	374	KILDARE, NOEL C.
*122	286000362	n/a	--	--	---	FOOT, DONNA D.

INSTRUCTOR'S GRADEBOOK REPORTS (Figure 4): Include (a) a list of the assessment titles, dates and weights; (b) an alphabetic listing of subjects' (e.g., students') names, their scores on every assessment to date, and the weighted total; and, (c) a rank ordered listing of subject names with their cumulative weighted total scores. Similar reports, one of which is shown in Figure 9, are available for a single assessment.

Figure 5. Gradebook Posting Reports

Z-Score GRADEBOOK POSTING REPORT: All Scores To Date Prepared 21-SEP-87 07:49

Z-SCORES

ID	Test:	1	2	3	4	Weighted	Total	Avg *
	Weight:	1	1	1	3	Total	Weights	
286000362		ABSENT	ABSENT	ABSENT	ABSENT	N/A	N/A	N/A
286001297		538	ABSENT	799	619	3194	5	638.80
286001645		535	ABSENT	280	497	2306	5	461.20
286002197		624	ABSENT	544	298	2062	5	412.40
286007980		ABSENT	580	721	558	2975	5	595.00

Z-Score GRADEBOOK POSTING REPORT: Cumulative Totals To Date Prepared 21-SEP-87 07:49

Z-SCORES

	ID	Avg	Rank	%ile	Final Z
* 24	286000362	N/A	--	--	---
43	286001297	638.80	4	96	701
44	286001645	461.20	93	23	445
77	286002197	412.40	110	9	374
123	286007980	595.00	12	90	638

GRADEBOOK POSTING REPORTS (Figure 5): Provide listings of subjects' (e.g., students') scores in order of their confidential ID numbers but omitting their names. One part of the report shows each individual's scores on all assessments to date. The other gives the cumulative weighted total expressed in several units of measurement. A similar report (not shown) for a single (current) assessment is also available.

RATING SHEET AND INVENTORY (Figure 6): Provide rating data collection and automated computer input. The computer's line printer over-prints both identifying information and the inventory text (i.e., scale definitions for 5-point scale, rating criteria: categories and or items) on the machine scannable sheets. This includes "slugging" identification data in the scannable data grids. This provides flexibility while also reducing the total amount of data that must be manually entered on the sheet. Written comments may be provided by the rater on either the back of the sheet or on the front in areas not used for the text of rating items. The subject's ID number used on the rating sheet is not confidential. A different, confidential ID number is used in OTS-PR reports and records.

While the basic design of the processing programs and the rating sheet assumes brief (i.e., 1 criterion per line; 35 characters per descriptor) items and category labels will be used, more extended definitions can be used in either of two ways. A separately produced assessment guide may be given to the rater along with the rating sheet. The guide provides more detailed information on the criteria briefly given on the rating sheet. The category function can be used to print multi-line definitions of criteria on the rating sheet. On the rating sheet a category (sub-scale) begins with a category name or label. On immediately successive lines, the category label is followed by all the specific items defining the category.

Up to 30 different inventories may be used within a single course. The same or different inventories may be used from course to course. Each rating sheet has both a

course ID and inventory ID number printed on it in scannable form. When the scanner is processing rating forms it ejects (into a separate bin) any that do not fit the specified IDs for the current job. This prevents data of different types or from different sources from being inadvertently mixed together. A more detailed description of the Performance Rating portion of the system and descriptions of its application to a residency and four junior year clerkships are available in an earlier report (Cason, Schoultz, Cason, Glenn, Jones, Golden, Lang, and Doyle, 1986)

Figure 6. Rating Sheet and Over-printed Generic Inventory

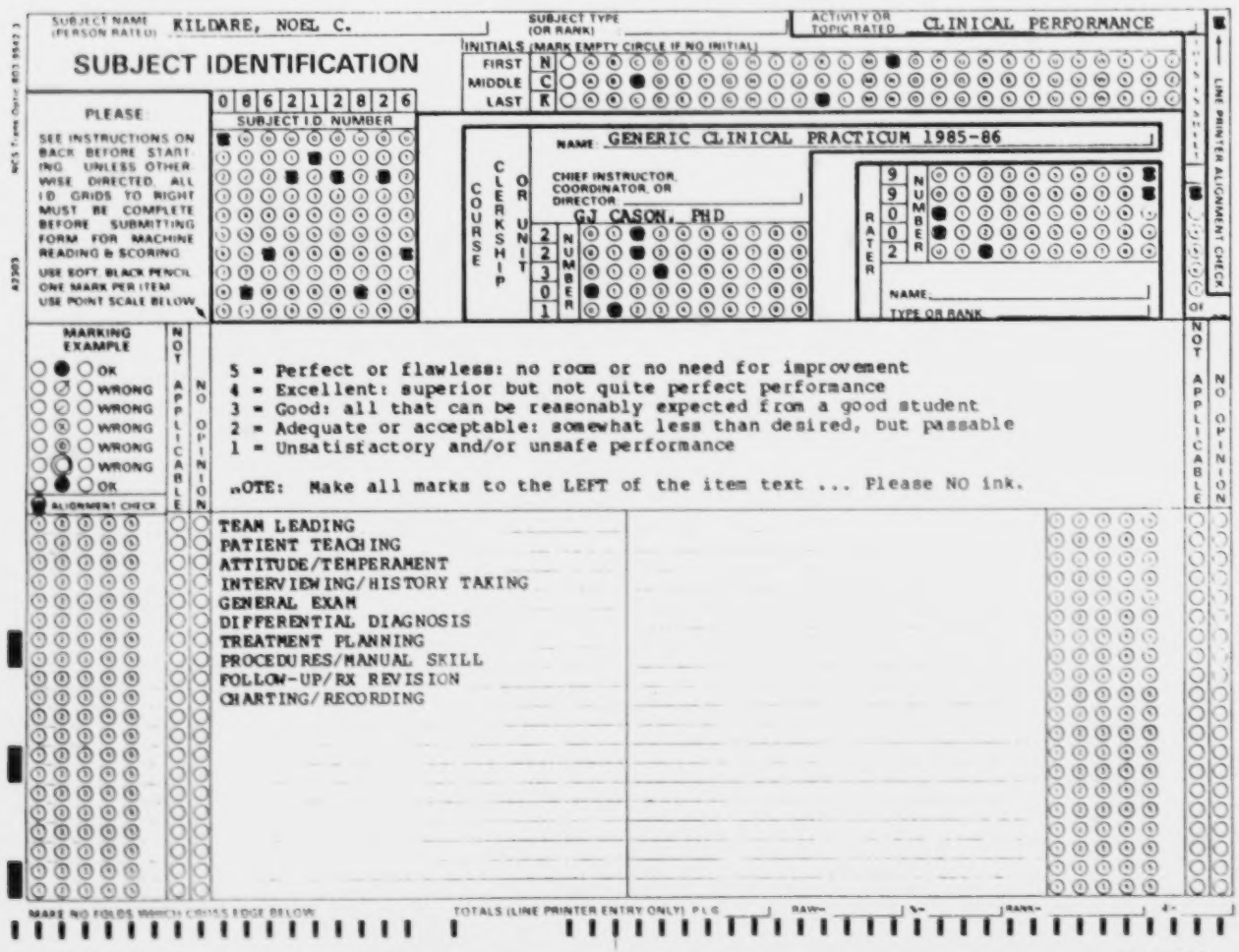

SUBJECT NAME (PERSON RATED) KILDARE, NOEL C.
SUBJECT TYPE (OR RANK)
ACTIVITY OR TOPIC RATED CLINICAL PERFORMANCE
SUBJECT IDENTIFICATION
INITIALS (MARK EMPTY CIRCLE IF NO INITIAL)
FIRST N
MIDDLE C
LAST K
0 8 6 2 1 2 8 2 6
SUBJECT I.D. NUMBER
PLEASE
SEE INSTRUCTIONS ON BACK BEFORE STARTING UNLESS OTHERWISE DIRECTED, ALL ID GRIDS TO RIGHT MUST BE COMPLETE BEFORE SUBMITTING FORM FOR MACHINE READING & SCORING
USE SOFT BLACK PENCIL ONE MARK PER ITEM USE POINT SCALE BELOW
COURSE OR CLERKSHIP UNIT NUMBER
NAME GENERIC CLINICAL PRACTICUM 1985-86
CHIEF INSTRUCTOR, COORDINATOR, OR DIRECTOR
GJ CASON, PHD
2 2 3 0 1
RATER NUMBER
9 9 0 0 2
NAME
TYPE OR RANK
MARKING EXAMPLE
OK
WRONG
WRONG
WRONG
WRONG
WRONG
OK
ALIGNMENT CHECK
NOT APPLICABLE
NO OPINION
5 = Perfect or flawless: no room or no need for improvement
4 = Excellent: superior but not quite perfect performance
3 = Good: all that can be reasonably expected from a good student
2 = Adequate or acceptable: somewhat less than desired, but passable
1 = Unsatisfactory and/or unsafe performance
NOTE: Make all marks to the LEFT of the item text ... Please NO ink.
TEAM LEADING
PATIENT TEACHING
ATTITUDE/TEMPERAMENT
INTERVIEWING/HISTORY TAKING
GENERAL EXAM
DIFFERENTIAL DIAGNOSIS
TREATMENT PLANNING
PROCEDURES/MANUAL SKILL
FOLLOW-UP/RX REVISION
CHARTING/RECORDING
MAKE NO FOLDS WHICH CROSS EDGE BELOW
TOTALS (LINE PRINTER ENTRY ONLY) PLG
RAW=
%=
RANK=
LINE PRINTER ALIGNMENT CHECK

Figure 7. Individual Performance Report with Generic Inventory

INDIVIDUAL PERFORMANCE REPORT (Current Rating)

Prepared 29-Jul-86 14:53 by the UAMS OTS/PR System (version B0) as implemented at UAMS

To: NOEL C. KILDARE
From: GJ CASON, PHD
Re: Rating/test 3-INPATIENT RATING - FACULTY

Dept: EDUCATIONAL DEVELOPMENT
Course: GENERIC CLINICAL PRACTICUM (SYNTHETIC DATA) 1985-86

Class Overall Mean Rating = 3.74-->
Your Overall Mean Rating = 3.92-->

Item	5 Point Scale 1.....2.....3.....4.....5	5 Pt Score — Your Mean=x	SEM	Class Mean=c	StdDev	Raw Score Perfect p	Yours xp/5	N of Raters
1 TEAM LEADING	.-.x	4.00	.32	3.97	.428	5	4.00	4
2 PATIENT TEACHING	.x.-	3.75	.36	3.80	.513	5	3.75	4
3 ATTITUDE/TEMPERMENT	.-.-c---x	4.75	.32	4.08	.455	5	4.75	4
4 INTERVIEWING/HISTORY TAKING	.-c---x	4.50	.38	3.92	.540	5	4.50	4
5 GENERAL EXAM	x---c-.-	3.00	.38	3.67	.552	5	3.00	2
6 DIFFERENTIAL DIAGNOSIS	x-c.-.-	3.33	.37	3.63	.481	5	3.33	3
7 TREATMENT PLANNING	xc.-.-	3.50	.30	3.61	.435	5	3.50	4
8 PROCEDURES/MANUAL SKILL	cx-.-	3.67	.41	3.64	.486	5	3.67	3
9 FOLLOW-UP/RX REVISION	xc.-.-	3.50	.39	3.64	.555	5	3.50	2
10 CHARTING/RECORDING	x---c.-.-	3.00	.39	3.57	.494	5	3.00	2

Rating Scale
5 = Perfect or flawless: no room or no need for improvement
4 = Excellent: superior but not quite perfect performance
3 = Good: all that can be reasonably expected from a good student
2 = Adequate or acceptable: somewhat less than desired, but passable
1 = Unsatisfactory and/or unsafe performance

Definition of Symbols
C = class overall 5 pt score: 3.74 StdDev: .398
X = your overall 5 pt score: 3.92 SEM : .24
c = class mean 5 pt score on item (or category)
x = your mean 5 pt score on item (or category)
SEM= Standard Error of Measurement

Your overall raw score 39.17 (out of perfect 50) yields: 78.3% Z= 544 Rank= 18 (out of 60). Class ave raw score 37.41

INDIVIDUAL PERFORMANCE REPORT (Figure 7): Provides information on one individual. Provides item level average ratings (across however many raters rated this individual) in both tabular and graphic form. The individual's scores ("x" profile in graph) and the averages of the comparison group (i.e., class, cohort; "c" profile in graph) are given. The standard error of measurement is computed taking into account the number of raters rating the particular subject. Individual rater reports of similar structure and appearance may be used for feedback to raters as suggested by Littlefield, Harrington, Anthracite, and Garman (1981). The rater's report gives the average rating that the rater assigned to subjects compared with the average of rater averages. Performance at the category (i.e., sub-scale) and total score level for all members of the group are reported in the Department Report (not shown). The Letter Report (not shown) provides information on a single individual's item-by-item performance on a multiple-choice exam: the examinee's response is shown for each item and the correct answer(s) is shown if different from the examinee's.

Figure 8. Individual Cumulative Score Report

INDIVIDUAL CUMULATIVE SCORE REPORT

Prepared 29-Jul-86 15:03 by the UAMS OTS/PR System (version B0) as implemented at UAMS

To: NOEL C. KILDARE
From: GJ CASON, PHD
Re: Your test scores and cumulative scores to date

Dept: EDUCATIONAL DEVELOPMENT
Course: GENERIC CLINICAL PRACTICUM (SYNTHETIC DATA) 1985-86

TEST NO.	GROUP MEAN RAW SCORE	GROUP SIZE	TEST DATE	TEST NAME	TEST WEIGHT	RAW SCORE	MAXIMUM POINTS	NORMALIZED TO PERCENT	NORMALIZED TO 5 POINTS	RANK	Z
1	38.72	58	19-Jun-86	OUTPATIENT RATING - SUPERVISOR	1	44.50	45.00	98.89	4.94	3	624
2	34.84	61	19-Jun-86	EMERGENCY ROOM RATING - STAFF	1	ABSENT	45.00				
3	37.41	60	29-Jul-86	INPATIENT RATING - FACULTY	1	39.17	50.00	78.33	3.92	18	544
4	50.30	121	29-Jul-86	COMPREHENSIVE EXAMINATION	3	30.50	100.00	30.50	1.53	116	298
	Currently Enrolled:128			Total applicable weights:	5						
				Your weighted normalized scores, and derived rank and Z:		———	———	53.74	2.69	105	403
				The weighted mean of your Z scores is 412.4, derived rank and Z:		———	———	———	———	110	374

INDIVIDUAL CUMULATIVE SCORE REPORT (Figure 8): Provides information to or about a single individual. Provides a list of all assessments (i.e., tests, ratings) to date, class averages, this individual's scores on each, and the weighted total computed in alternative ways and expressed in several units of measurement. Cumulative information on all subjects in the group is provided in the various Gradebook Reports (Figures 4 & 5).

Figure 9. Students by Rank Report

STUDENTS BY RANK (Current Rating) Prepared 30-Jul-86 14:27 by the UAMS OTS/PR System (version B0) as implemented at UAMS

Test: 3 - INPATIENT RATING - FACULTY
Instructor: GJ CASON, PHD
Course: GENERIC CLINICAL PRACTICUM (SYNTHETIC DATA) 1985-86

Dept: EDUCATIONAL DEVELOPMENT
Slot: 595 Phone:661-5720
Subjects rated: 60 Absent: 68 Withdrawn: 8

		RAW SCORE	5 PT SCORE	RANK RAW	RANK PCNTL	Z SCORE	N OF RATERS	NAME
	Averages	37.41	3.74			500	3.8	(Fictitious names and data)
1	286001297	49.33	4.93	1	98	799	3	ZEIMANOVICH, MARINA K.
2	286007980	46.22	4.62	2	96	721	3	BEAKER, JULIO M.
18	286002197	39.17	3.92	18	70	544	4	KILDARE, NOEL C.
60	286001645	28.67	2.87	60	0	281	3	ARNOLD, FRANKLIN D.

STUDENTS BY RANK REPORT (Figure 9): Provides information about all the members of a group. Provides a rank ordered listing of students' names, confidential ID numbers, scores in various units of measure, and number of raters rating each student. When rating sheets (Figure 6.) are used to collect performance rating data and raters are given unique ID numbers, a similar Raters by Rank Order Report is produced. It gives a listing of all raters from the rater that gave the highest average rating to the rater that gave the lowest. Although OTS-PR currently has no provision for automatically off-setting differences in raters' stringency (Cason and Cason, 1984, 1985; Delk, Cason and Reese, 1986), one of the uses of the rater reports is the detection of such differences.

Figure 10. Rating Analysis Summary

RATING ANALYSIS SUMMARY (Current Rating) Prepared 29-Jul-86 14:40 by the UAMS OTS/PR System (version B0) as implemented at UAMS

Test: 3 - INPATIENT RATING - FACULTY
Instructor: GJ CASON, PHD
Course: GENERIC CLINICAL PRACTICUM (SYNTHETIC DATA) 1985-86

Dept: EDUCATIONAL DEVELOPMENT
Slot: 595 Phone:661-5720
Subjects rated: 60 Absent: 68 Withdrawn: 8

	Category	N Rated Items	Total Points (1)	Average Score Raw	5 Pt	1 Rater Reliab.	Mean N Raters & Reliability (2)		Standard Dev Raw	5 Pt	Std Error of Measure Raw	5 Pt	2 (3)
1	TEAM LEADING	1	5	4.0	3.97	0.17	3.7	0.43	0.4	.428	0.32	.32	75
2	PATIENT TEACHING	1	5	3.8	3.80	0.21	3.7	0.49	0.5	.513	0.37	.37	71
3	ATTITUDE/TEMPERMENT	1	5	4.1	4.08	0.20	3.7	0.48	0.5	.455	0.33	.33	72
4	INTERVIEWING/HISTORY TAKING	1	5	3.9	3.92	0.21	3.5	0.48	0.5	.540	0.39	.39	72
5	GENERAL EXAM	1	5	3.7	3.67	0.36	2.7	0.60	0.6	.552	0.35	.35	63
6	DIFFERENTIAL DIAGNOSIS	1	5	3.6	3.63	0.18	3.2	0.41	0.5	.481	0.37	.37	77
7	TREATMENT PLANNING	1	5	3.6	3.61	0.21	3.6	0.49	0.4	.435	0.31	.31	72
8	PROCEDURES/MANUAL SKILL	1	5	3.6	3.64	0.12	2.1	0.22	0.5	.486	0.43	.43	88
9	FOLLOW-UP/RX REVISION	1	5	3.6	3.64	0.35	2.8	0.60	0.6	.555	0.35	.35	63
10	CHARTING/RECORDING	1	5	3.6	3.57	0.24	3.0	0.49	0.5	.494	0.35	.35	72
	Overall	10	50	37.4	3.74	0.30	3.8	0.61	4.0	.398	2.47	.25	62

RATING ANALYSIS SUMMARY REPORT (Figure 10): Provides information on the performance of the rating inventory and procedure. Information is given at the category (sub-scale) and total score level. If two or more raters rate each subject, meaningful inter-rater reliabilities (intra-class correlations) are computed. Reliabilities are reported for a single rater and for the average number of raters rating a subject. The second of these provides a measure of the reliability of the current assessment procedure. Were no changes made in the inventory or procedure, the Spearman-Brown expansion formula could be applied to the single rater reliability to determine the number of raters (rating each subject) needed for the average of their ratings to attain a desired level of reliability. More detailed information, such as the distribution of responses on each item and performance of items-within-categories, is provided in the Rating Item Analysis Report (not shown). When a multiple-choice exam is scored by OTS-PR, a Test Item Analysis Report and a Test Analysis Summary are always automatically produced. These reports

(not shown) are similar in structure, content, and purpose to the Rating Analysis reports. The information contained in these reports can assist in the improvement of assessment procedures. As suggested in Figure 1, at UAMS a member of the staff of the Office of Educational Development is available to assist faculty to make best use of the information in assessment analysis reports.

ALTERNATE TEST FORMS: Provide a means of greatly reducing the probability of one examinee copying answers from another. Up to five different orderings of the same questions, i.e., five forms of the same test, may be created and processed as a single test by OTS-PR. One arrangement of the questions is designated as Form 1, then a special scan sheet (not shown) is used to define the mapping of questions in the other forms relative to Form 1. A special report on the means and standard deviations of each test form separately and for all data collectively is produced. This permits checking to see that the test forms are parallel and equivalent, that is that changing the order of questions did not result in substantially different tests. Other than indicating the test form a particular examinee took, most of the other reports remain essentially the same as when a single test form is used. Test and item analyses are based on all data from all forms as are examinees' group statistics (mean, SD, Z-scores). The Letter Report is the exception because it provides item by item information to an examinee. When multiple forms are used, the Letter report indicates which form an examinee had, the correct answer(s) if different from that given, and the corresponding item number on Form 1. The last simplifies post-test discussion and records keeping.

Results of OTS-PR System Implementation

In an average month OTS-PR currently supports about 50 courses, clerkships, and other applications in the four UAMS colleges. This is approximately the same as the number of courses that are offered on average. The majority of UAMS courses make some use of the test scoring and or performance rating resources available through OTS-PR. Over the course of a year, about 60,000 test answer sheets and about 60,000 rating sheets are used to support evaluation of both students and teaching performance of faculty at UAMS. Although the PR portion of the system was originally developed to assist the clinical teacher to evaluate the clinical performance of students (and that remains its primary use), it has also proven useful in the evaluation of faculty (Coleman, Thompson & Cason, 1986). The PR portion of the system has also been used in a study of the systematic variation in the stringency of proposal reviewers for several annual meetings of a research organization (Cason, Cason & Stritter, 1986a; 1986b) and as part of the process to actually off-set such rater differences prior to the selection of papers for an international clinical research congress (Cason, Cason & Redland, 1987).

In the College of Medicine, all the Basic Science courses use OTS to evaluate student performance. Faculty are encouraged to view the scoring based on the initial answer key as much like the first draft of an article: almost certainly in need of revision (Cason, 1978). In most of the courses, information from the technical item analyses provided by OTS, student appeals of the correct answer(s) used in the initial scoring key, and or post-test discussion sessions lead to re-scoring of the tests with revised keys. This process is facilitated by the item types designed for re-scoring purposes. The resulting tests tend to have higher reliability and they are seen as more valid by faculty and students alike. Also, contributing to the improvement in tests are a gradual shift from use of NBME type K items to NBME type X and more essay questions. Almost all courses using OTS for test scoring have students' ID data pre-printed on the answer sheets by the computer. This has virtually eliminated problems with students not remembering or improperly entering this information and the consequent delay in obtaining complete reports and wasted computer time.

In 1981, when the PR portion of the system first became operational, only one clerkship used the system (Sub-specialties). Now, two residencies (Anesthesia and Pediatrics) and all the Junior Year clerkships except Medicine and Psychiatry (which use manual data entry and micro-computers) have chosen to use OTS-PR to collect and process student clinical performance data. Medicine's and Psychiatry's failure to adopt use of OTS-PR is somewhat ironic since the prototype system from which the PR portion of OTS-PR evolved was field tested in the Medicine and Psychiatry clerkships in 1978-79 and 1979-80. Even though that prototype was relatively cumbersome and far less informative, it was primarily upon the recommendation of the faculty in these clerkships, based on their experience with the prototype system, that the resources were invested to develop and implement the fully operation Performance Rating enhancement of OTS (Cason, 1981). Initially, this situation probably resulted in part from the 18 month interruption in rating information processing support between discontinuation of the use of the prototype programs and implementation of the first version of the PR programs in October of 1981. Part of the reason it may continue is that these clerkships have adopted a procedure (Delk, Cason & Reese, 1986), implemented via micro-computer, to off-set differences in rater stringency. OTS-PR at present does not offer this capacity. While a combined approach offering the advantages of each is possible, they have not chosen to use it.

In all our clerkships, as recommended by Ramsey, Shannon, Flemming, Wenrich, Peckham, and Dale (1986), both objectively scored exams and clinical performance ratings are used in the assessment of students. Thus, the clerkships tend to make broadest use of the range of capacities of the OTS-PR system. The flexibility of the system has permitted each clerkship to develop one or more rating inventories that suit its own teaching goals and clinical circumstances. The number of independent ratings of students' clinical performance in our clerkships vary from 2 in Psychiatry and Surgery to 18 in Pediatrics. The number of different rating inventories used in a single clerkship varies from 1 to 4. Among the strengths of OTS-PR cited by clerkship directors that use it are: ease of use; flexibility, i.e., being able to have inventories tailored to each clerkship's specific needs; getting more uniform reports from different training sites; the usefulness of the reports, particularly the individual performance reports on both students and raters (faculty, residents) with their graphic display of individual versus group performance on each criterion; and, avoiding problems many see as inherent in grading sessions (various biasing effects, scheduling, etc). Chairs of clinical departments have reported that the information obtained through the OTS-PR system is of far superior quality to that given in narrative statements prepared by faculty. This perception arises in part because the OTS-PR individual performance report gives both relative strengths and weaknesses of a student. The narrative statements prepared by faculty tend to avoid any negative comments.

Conclusions

As it stands, OTS-PR goes a long way toward making it easy for faculty to use an appropriate assessment approach although there are still many enhancements to the system needed, for example, automatic correction for differences in raters' stringency in clinical courses, estimation of the reliability of a cumulative score based on several assessments (tests and or ratings), automatic generation of narrative performance reports for use in preparation of "Dean's Letters", and so forth. The results of faculty access to this system suggest that when faculty are given student performance assessment tools, such as OTS-PR, that make assessment relatively easy and convenient, the quality control of medical education improves. More attention needs to be focused on developing and implementing practical, "user friendly" systems to support the newer, more sophisticated, performance assessment methodologies coming out of medical education research efforts. These methodologies require flexible data management, analysis, and reporting systems if they are to move from the research study to everyday educational use. In 1973, Printen, Chappell and Whitney described a system for processing both exam and rating data that

appears to have been about as cumbersome as our earliest prototype system in 1978. Although, they reported at that time that they expected their system to be adopted in other clerkships in their school, a recent (December, 1986) inquiry indicated that it was still used only in the surgery clerkship. If better assessment methods are to be adopted, exhortation and technical excellence of the methods are far less important than making it easy for medical educators to use them in actual day to day educational practice.

References

Albanese, M.A., Kent, T.H., & Whitney, D.R. (1979). Cluing in multiple-choice test items with combinations of correct responses. *Journal of Medical Education*, *54*(12), 948-50.

Cason, C.L., Cason, G.J., & Redland, A. (1987). Off-setting differences in reviewer stringency. Paper presented at the International Research Congress, Edinburgh, Scotland, sponsored by Sigma Theta Tau, International Honor Society of Nursing, Indianapolis, Indiana, USA; the Royal College of Nursing, London, England, UK; and, the Department of Nursing Studies, University of Edinburgh, Edinburgh, Scotland, UK, July.

Cason, C.L., Cason, G.J., & Stritter, F.T. (1986a). Reviewer standards in Division I program selection. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, April. Abstract in *Resources in Education*, *21*(10), 145. ERIC Document ED-270-458.

Cason, C.L., Cason, G.J., & Stritter, F.T. (1986b). Reviewer stringency and proposal quality in the selection of the 1986 AERA Division I program. *Professions Education Research Notes*, *8*(1), 7-8.

Cason, G.J. (1978). **Guidelines for the Objective Testing of Students**. Little Rock, Arkansas: Office of Educational Development, University of Arkansas for Medical Sciences. Abstract in *Resources in Education*, 1981, 16(8) 194. ERIC Document ED-200-628.

Cason, G.J. (1986). Item Type K versus X. *Professions Education Research Notes*, *8*(1), 3. (Letter)

Cason, G.J. & Cason, C.L. (1984). A deterministic theory of clinical performance rating: Promising early results. *Evaluation & the Health Professions*, *7*(2), 221-247.

Cason, G.J. & Cason, C.L. (1985). A regression solution to Cason and Cason's model of clinical performance rating: Easier, cheaper, faster. Paper presented at the annual meeting of the American Educational Research Association, Chicago, April. Abstract in *Resources in Education*, 1986, *21*(2), 147. ERIC Document ED-262-079.

Cason, G.J., Schoultz, T.W., Cason, C.L., Glenn, R.E., Jones, J.G., Golden, K.A., Lang, N.P., & Doyle, K.L. (1986). A flexible system for processing clinical performance ratings: Illustrative applications in a residency and four clerkships. Paper presented as part of an exhibit in Innovation in Medical Education at the 97th annual meeting of the Association of American Medical Colleges, New Orleans, Louisiana, October 1986. Abstract in *Resources in Education*, 1987, *22*(6), 144. ERIC Document ED-278-692.

Coleman, E.A., Thompson, P.J., & Cason, G.J. (1986). The role of centralized computer services in faculty evaluation. *Computers in Nursing*, November/December, 259-264.

Delk, J.L., Cason, G.J., & Reese, W.G. (1986). A method to enhance the fairness of clerkship ratings: A practical application. *Journal of Medical Education*, *60*(12), 944-45.

Harasym, P.H., Norris, D.A., & Lordschieder, F.L. (1980). Evaluating student multiple-choice responses. *Evaluation & the Health Professions*, *3*, 63-84.

Hubbard, J.P. (1971). **Measuring Medical Education: The Test and the Test Procedures of the National Board of Medical Examiners**. Philadelphia: Lea and Finbiger, 20-31.

Littlefield, J.H., Harrington, J.T., Anthracite, N.E., & Garman, R.E. (1981). Description and four year analysis of a clinical clerkship evaluation system. *Journal of Medical Education*, *56*, 334-340.

Printen, K.J., Chappell, W., & Whitney, D.R. (1973). Clinical performance evaluation of junior medical students. *Journal of Medical Education*, *48*, April, 343-348.

Ramsey, P.G., Shannon, N.F., Flemming, L., Wenrich, M., Peckham, P.D., and Dale, D.C. (1986). Use of objective examinations in medicine clerkships: Ten-year experience. *The American Journal of Medicine*, *81*, October, 669-674.

Whitley, T.W., Jones, J.G., & Revicki, D.A. (1986). Alternatives to multiple true-false items. *Professions Education Research Notes*, *7*(4), 5-6.

Acknowledgements

The successful day to day operation of the OTS-PR system is a reflection of the consistently friendly manner and efficient work of Mark Boughter, Academic Support Center Manager, and Faye Billings, Assessment Scoring Coordinator. Systems Analyst Tom Lewis programmed first the experimental prototype rating processing software then the first operational PR system programs. Since 1981, Academic Systems Manager Tom Hart and Programmers Chuck Desjardin and Jalene Hawk have programmed the many revisions and enhancements of the OTS-PR system; including the conversion of the system from DEC-10 to VAX-8530 main-frame in 1987.